***One dataset, many users.***

***Ten use cases and motivations for building a machine learning model using the Lending club dataset.***

**Performed by:** Team 15

**Task 1**

1. **Research and Summarize the Client Needs**
2. **Rick the Investor ( Risk averse )**

* It is an investor who when faced with two investments of similar expected returns, chooses to invest in the one with lower risk.

1. **Tola the Investor ( Risk taker )**

* It is an investor who is willing to take on additional risk for an investment that has a relatively low additional expected return in exchange for that risk.
* The investments here are prone to very high or very low returns , i.e. Very high kurtosis in a return distribution.

1. **Taz the borrower ( Good credit )**

* Indicates that the borrower has a relatively high credit score and is a safe credit risk.
* A good credit score depends on amount owed overall, length of credit history, types of credits used, new credits and credit inquiries.
* A credit score of 580 - 669 is considered as a fair credit score, and for above 670 it is good.

1. **Pip the borrower ( Bad credit )**

* Predicted inability to repay a debt on time and in full.
* This makes it difficult to get approved for a new loan, or at least will be offered credit only at a very high rate of interest.
* A credit score of 579 and lower is considered as a bad credit score.

1. **Bipa the Portfolio Manager**

* Person responsible for investing funds into assets and implementing its investment strategy and managing day to day portfolio trading.

1. **Arc the Arbitrager**

* Arbitrage is the simultaneous purchase and sale of an asset in order to profit from a difference in price
* Loaning a certain amount from the Lending Club on a certain interest rate and then giving that amount to borrowers on a certain interest rate.
* **Disadvantages:**

1. A substantial chunk of earnings from arbitrage will be taken away by taxes
2. Doesn’t offer a greater return
3. It is a high-risk investment trading strategy
4. **Slick the Data Scientist**

* Will be responsible to provide meaningful information based on large amounts of complex data.
* Combines different fields of work in statistics and computation in order to interpret data for the purpose of decision making.

1. **Irs the Professor**

* Would educate the next generation of clients about this analysis.

1. **Dat the data vendor who sells data and insights**

* Provides market data to financial firms, traders and investors. The data distributed is collected from sources such as stock exchange feeds, brokers and dealer desks or regulatory fillings.

1. **Mar the regulator**

* regulates how the model can/cannot be used

1. **Explore the data and comment on data quality, features and get a feel for the data**

* We have 74 columns of data out of which 20 are Categorical Variables and 54 are Numeric Variables.
* 38 of 74 columns allow NULL values.
* Most Categorical Variables are Text (Descriptions) whereas the others are Date (Days of Week, Months).
* Some Numeric Variables are derived based on the values of Other Numeric Variables.
* **Borrower Features:** member\_id, emp\_title, emp\_length, home\_ownership, annual\_inc, zip\_code, addr\_state, dti, delinq\_2yrs, earliest\_cr\_line, inq\_last\_6mths, mths\_since\_last\_delinq, mths\_since\_last\_record, open\_acc, pub\_rec, revol\_bal, revol\_util, total\_acc, acc\_now\_delinq, tot\_cur\_bal, open\_acc\_6m, open\_il\_6m, open\_il\_12m, open\_il\_24m, mths\_since\_rcnt\_il, total\_bal\_il, il\_util, open\_rv\_12m, open\_rv\_24m, max\_bal\_bc, all\_util, total\_rev\_hi\_lim, inq\_fi, total\_cu\_tl, inq\_last\_12m
* **Loan related Features:** id, loan\_amnt, funded\_amnt, funded\_amnt\_inv, term, int\_rate, installment, grade, sub\_grade, verification\_status, issue\_d, loan\_status, pymnt\_plan, url, desc, purpose, title, initial\_list\_status, out\_prncp, out\_prncp\_inv, total\_pymnt, total\_pymnt\_inv, total\_rec\_prncp, total\_rec\_int ,total\_rec\_late\_fee ,recoveries ,collection\_recovery\_fee ,last\_pymnt\_d ,last\_pymnt\_amnt ,next\_pymnt\_d ,last\_credit\_pull\_d ,collections\_12\_mths\_ex\_med ,mths\_since\_last\_major\_derog, ,policy\_code, application\_type, annual\_inc\_joint, dti\_joint, verification\_status\_joint, tot\_coll\_amt

1. **Key takeaways and insights from charts for the clients**
2. Link 1 : <https://www.lendingclub.com/info/statistics.action>

* From the total loan issuance graph, it can be seen that the amount of loans disbursed year over year increased as we go from 2011 to 2018, and within the years, the growth is exponential in every quarter.
* The maximum reported loan purpose radial pie shows that of Refinancing contributed a total of 45.6% followed by Credit card payoffs which held a total of 22.58% and then the remaining Other 31.82%
* Loan issuance by state graph tells us that over 19 states in the US issued total loan amounts of over $50 million and more, whereas 14 states held a total of $25-50 million and the remaining 18 states issued total loan amount between $0-25 million.

1. Link 2: <https://www.lendingclub.com/info/statistics-performance.action>

* Investor account returns by average age of portfolio shows that of all investor accounts on the LendingClub platform that have invested in at least 100 Notes and that have not purchased or sold Notes on the Folio Investing Note Trading Platform.
* We have adjusted NAR on the Y axis against average age of portfolios. It is the weighted average age of notes.
* If I invest in multiple loans then my adjusted NAR would be high in the beginning and over the period of time as we traverse, the adjusted NAR decreases due to multiple reasons, one of which is the loan can end up being charged off.
* If a loan is concentrated and that loan ends up being charged off, then the NAR would be 0, therefore it is advisable to invest into multiple loans with at least the minimum amount of Notes rather than just investing into one loan with all your funds.
* Over an increase in time, the adjusted NAR falls.
* Also, it can be seen that diversification can reduce the volatility of returns. So, an increase in the number of notes reduces volatility. For example, 100+ notes, so the more number of notes, the less risk.

1. Link 3: <https://www.lendingclub.com/info/demand-and-credit-profile.action>

* Average interest rate multi-line chart shows the comparison of Loan Grades and their Terms over the years v/s the interest rates.
* As we move from grades A to G, it can be seen that there is an increase in the interest rate of those loans. It also means that a loan with a grade of A will yield a lower interest rate as compared to that of a G grade loan.
* Over the years, it can be seen that the loan grades A to D have a constant interest rate band whereas there is a significant increase in the interest rates with loan grades E to G with a substantial increase of 10%
* Grade mix over time shows that lower the grade of the loan, lower issuance for that loan have been made over the years. Loan grades B, C followed by A were the maximum issued loan grades.
* Over time, the amount of loans taken by borrowers have been repaid in a timely manner without the loan going default. It can be seen in the Net annualized returns chart.
* Loan status migration over 9 months shows us that the loans which are past due, 22% of those loans end up being in the net-charged off status in the grace period whereas 78% of the loans in the grace period remain active. As we move ahead of the timeframe, we can see that most of the loans end up being net-charged off if the duration is significantly over the grace period.